Sharp Bounds on the Entropy of the Poisson Law and Related Quantities

José A. Adell, Alberto Lekuona, and Yaming Yu, *Member, IEEE*

Abstract—One of the difficulties in calculating the capacity of certain Poisson channels is that $H(\lambda)$, the entropy of the Poisson distribution with mean λ, is not available in a simple form. In this work we derive upper and lower bounds for $H(\lambda)$ that are asymptotically tight and easy to compute. The derivation of such bounds involves only simple probabilistic and analytic tools. This complements the asymptotic expansions of Knessl (1998), Jacquet and Szpankowski (1999), and Flajolet (1999). The same method yields tight bounds on the relative entropy $D(n, p)$ between a binomial and a Poisson, thus refining the work of Harremoës and Ruzankin (2004). Bounds on the entropy of the binomial also follow easily.

Index Terms—asymptotic expansion, binomial distribution, central moments, complete monotonicity, entropy bounds, integral representation, Poisson channel, Poisson distribution

I. INTRODUCTION

Unlike the differential entropy for the Gaussian distribution, the Shannon entropies for many basic discrete distributions, such as the Poisson, the binomial, or the negative binomial, are "not in closed form." In the Poisson case, the lack of a simple analytic expression is seen ([24], [22]) as one of the obstacles to obtaining the capacity of certain Poisson channels ([6], [12], [26], [27]). Computation of the entropy is also a basic problem partly motivated by the maximum entropy characterizations ([28], [25], [14], [29], [18], [30]) of these distributions.

One strategy to make the entropy functions more tractable is to express them as integrals ([10], [20], [24]). See [13], [21], [23] for related integral representations for entropy-like quantities in the context of Poisson channels. For the entropy functions themselves, integral representations have been used to derive asymptotic expansions ([10], [20]). Alternatively, asymptotic expansions can be obtained using analytic depoissonisation ([16], [17]), singularity analysis ([11]), or local limit theorems ([9]).

It is obviously desirable to have bounds that accompany asymptotic expansions, for both theoretical analysis and numerical computation. This is especially true for quantities such as the entropy of the Poisson law, which may be used, e.g., in capacity calculations for discrete-time Poisson channels ([24], [22]). Part of this work aims to derive tight bounds on the entropy for fundamental distributions such as the Poisson

José A. Adell is with the Departamento de Métodos Estadísticos, Facultad de Ciencias, Universidad de Zaragoza, Pedro Cerbuna 12, 50009 Zaragoza, Spain (email: adell@unizar.es).

Alberto Lekuona is with the Departamento de Métodos Estadísticos, Facultad de Ciencias, Universidad de Zaragoza, Pedro Cerbuna 12, 50009 Zaragoza, Spain (email: lekuona@unizar.es).

Yaming Yu is with the Department of Statistics, University of California, Irvine, CA, 92697-1250, USA (e-mail: yamingy@uci.edu).

and the binomial. Our results are expressed in terms of two sequences of lower and upper bounds, and have the following features.

- The sequence of upper bounds and the sequence of lower bounds each gives a full asymptotic expansion for the entropy. In other words the bounds are asymptotically tight.
- The bounds are derived from familiar quantities such as the central moments of the Poisson, and are in a form simple enough for both theoretical analysis and numerical computation.
- The derivation, which involves only real analysis, is elementary. (Note that the asymptotic expansions of Knessl [20] are also real-analytic.)

Denote by $\mathbf{Z}_+ = \{0, 1, 2, \ldots\}$ and by $\mathbf{N} = \mathbf{Z}_+ \setminus \{0\}$. As usual, the Shannon entropy for a discrete random variable X on $\mathbf{Z}_+$ with mass function $f_i = \Pr(X = i)$ is defined as

$$H(X) = H(f) = \sum_{i=0}^{\infty} -f_i \log f_i,$$

where we use the natural logarithm and obey the convention $0 \log 0 = 0$. Throughout we use N_λ to denote a Poisson random variable with mean λ, i.e., the mass function is

$$\Pr(N_\lambda = j) = \frac{e^{-\lambda}\lambda^j}{j!}, \quad j \in \mathbf{Z}_+.$$

We write $H(\lambda) = H(N_\lambda)$ for simplicity. The best known bound on $H(\lambda)$ is perhaps

$$H(\lambda) \le \frac{1}{2} \log\left(2\pi e\left(\lambda + \frac{1}{12}\right)\right), \tag{1}$$

which is obtained by bounding the differential entropy of $N_\lambda + U$ where U is an independent random variable uniformly distributed on $(0, 1)$ ([8], Theorem 8.6.5). While (1) is simple in form and reasonably accurate, it lacks a corresponding lower bound, and does not extend easily to capture higher order terms in the expansion of $H(\lambda)$. As a remedy we shall derive, for each $m \ge 1$, a double inequality of the form

$$\sum_{k=m}^{2m} \frac{A(m,k)}{\lambda^k} \le H(\lambda) - \frac{1}{2}\log(2\pi\lambda) - \frac{1}{2} - \sum_{k=1}^{m-1}\frac{b(m,k)}{\lambda^k}$$
$$\le \sum_{k=m}^{2m-1} \frac{B(m,k)}{\lambda^k}, \tag{2}$$

where $A(m, k)$, $b(m, k)$, and $B(m, k)$ are explicit constants ($\sum_1^0 \equiv 0$). In other words, for each $m \ge 1$, we give a finite

asymptotic expansion in powers of λ^{-1} with m exact terms and explicit lower and upper bounds of the order of λ^{-m}.

In Section II we derive (in an equivalent form) the double inequality (2). The key steps are

- an integral representation that relates $H(\lambda)$ to the simpler quantity $E[\log(N_\lambda + 1)]$;
- bounds on $E[\log(N_\lambda + 1)]$ in terms of polynomials in λ^{-1}, which translate easily to bounds on $H(\lambda)$.

Note that (2) is only effective for large λ. We also obtain bounds on $H(\lambda)$ in terms of polynomials in λ, which work for small λ.

Besides $H(\lambda)$, we also consider bounds on the relative entropy between a binomial and a Poisson, thus obtaining a version of "the law of small numbers" that refines the results of Harremoës and Ruzankin [15]. While these are of theoretical interest, they also lead to new bounds for the entropy of the binomial. As usual, for two random variables X and Y on $\mathbf{Z}_+$ with mass functions f and g respectively, the relative entropy is defined as

$$D(X\|Y) = D(f\|g) = \sum_{j=0}^{\infty} f_j \log \frac{f_j}{g_j}.$$

By convention $0\log(0/0) = 0$, and $D(f\|g) = \infty$ if f assigns mass outside of the support of g. Throughout we let $B_{n,p}$ be a binomial random variable with mass function

$$\Pr(B_{n,p} = k) = \binom{n}{k} p^k q^{n-k}, \quad k = 0, 1, \ldots, n,$$

where $q \equiv 1 - p$, $p \in (0, 1)$ and $n \in \mathbf{N}$. We consider

$$D(n, p) = D(B_{n,p}\|N_{np}),$$

i.e., the relative entropy between $B_{n,p}$ and N_{np}, and derive bounds on $D(n, p)$ using similar techniques. Bounds on the entropy of the binomial $H(B_{n,p})$ are obtained as a corollary.

Sections III and IV contain proofs of the main results. We conclude with a short discussion on possible extensions in Section V.

II. MAIN RESULTS

A. Sharp Bounds on $H(\lambda)$

We first present a class of double inequalities for $H(\lambda)$ that is effective for small λ (say $\lambda \leq 1$), but valid for all $\lambda \geq 0$.

Theorem 1: For any $\lambda \geq 0$ and $m = 1, 2, \ldots$, we have

$$\sum_{k=2}^{2m+1} \frac{c(k)}{k!} \lambda^k \leq H(\lambda) + \lambda \log \lambda - \lambda \leq \sum_{k=2}^{2m} \frac{c(k)}{k!} \lambda^k,$$

where

$$c(k) = \sum_{j=0}^{k-1} (-1)^{k-1-j} \binom{k-1}{j} \log(j+1), \quad k = 2, 3, \ldots.$$

For fixed m, the two bounds given by Theorem 1 differ by $O(\lambda^{2m+1})$. Hence they are most effective when λ is small. Moreover, inspection of the proof (Section III, Part B) shows that, for $0 \leq \lambda \leq 1$, both the upper and lower bounds in Theorem 1 converge to $H(\lambda) + \lambda \log \lambda - \lambda$ as $m \to \infty$.

In what follows, the kth central moment of the Poisson distribution

$$\mu_k(s) \equiv E[(N_s - s)^k]$$

plays an important role. The first few values of $\mu_k(s)$ are $\mu_0(s) = 1$, $\mu_1(s) = 0$, and

$$\mu_2(s) = \mu_3(s) = s, \ \mu_4(s) = 3s^2 + s, \ \mu_5(s) = 10s^2 + s.$$

They obey the well-known recursion ([19], p. 162)

$$\mu_k(s) = s \sum_{j=0}^{k-2} \binom{k-1}{j} \mu_j(s), \quad k \geq 2, \tag{3}$$

from which it is easy to show that, for $k \geq 2$, $\mu_k(s)$ is a polynomial in s of degree $\lfloor k/2 \rfloor$, where $\lfloor x \rfloor$ denotes the integer part of x.

In contrast to Theorem 1, Theorem 2 is most effective for large λ.

Theorem 2: For any $\lambda > 0$ and $m = 1, 2, \ldots$, we have

$$-r_m(\lambda) \leq H(\lambda) - \frac{1}{2}\log(2\pi\lambda) - \frac{1}{2} - \beta_m(\lambda) \leq 0,$$

where

$$\beta_m(\lambda) = \int_\lambda^\infty \left(\sum_{j=3}^{2m+1} \frac{(-1)^{j-1}\mu_j(s)}{j(j-1)s^j} \right) ds = \sum_{k=1}^{2m-1} \frac{b(m,k)}{\lambda^k} \tag{4}$$

and

$$r_m(\lambda) = \int_\lambda^\infty \frac{\mu_{2m+2}(s)}{(2m+1)s^{2m+2}} ds = \sum_{k=m}^{2m} \frac{a(m,k)}{\lambda^k}. \tag{5}$$

Let us note that the seemingly cumbersome expressions (4) and (5) are actually quite easy to handle. For $j \geq 3$, the jth central moment $\mu_j(s)$ is a polynomial in s of degree $\lfloor j/2 \rfloor$. This, together with (3), shows that the integrand in (4) is a polynomial in s^{-1}, with powers going from s^{-2} to s^{-2m}. Similar statements hold for the integrand in (5). Hence the constants $a(m,k)$ and $b(m,k)$ are obtained after straightforward integration; see Table I for their values for small m. In particular, for $m = 2$ we have

$$-\frac{31}{24\lambda^2} - \frac{33}{20\lambda^3} - \frac{1}{20\lambda^4} \leq H(\lambda) - \frac{1}{2}\log(2\pi\lambda) - \frac{1}{2} + \frac{1}{12\lambda}$$
$$\leq \frac{5}{24\lambda^2} + \frac{1}{60\lambda^3}.$$

We emphasize that the constants $b(m,k)$, $1 \leq k \leq m-1$, are exact in the full asymptotic expansion of $H(\lambda)$, since (5) gives $r_m(\lambda) = O(\lambda^{-m})$. For example, we have $b(3,1) = -1/12$ and $b(3,2) = -1/24$ from Table I, and hence

$$H(\lambda) = \frac{1}{2}\log(2\pi\lambda) + \frac{1}{2} - \frac{1}{12\lambda} - \frac{1}{24\lambda^2} + O(\lambda^{-3}),$$

which agrees with the leading terms given by, e.g., Knessl ([20], Theorem 2).

Bounds on $H(\lambda)$ given by Theorem 2 are illustrated in Fig. 1. As λ increases from 10 to 20, the gap between upper and lower bounds, $r_m(\lambda)$, decreases

- from ≈ 0.1 to ≈ 0.05 with $m = 1$,
- from ≈ 0.017 to ≈ 0.004 with $m = 2$, and

TABLE I
VALUES OF $a(m,k)$ AND $b(m,k)$ FOR $m = 1,2,3,4$.

k	$a(m,k)$			
	$m=1$	$m=2$	$m=3$	$m=4$
1	1			
2	1/6	3/2		
3		5/3	5	
4		1/20	35/2	105/4
5			17/5	210
6			1/42	2275/18
7				167/21
8				1/72

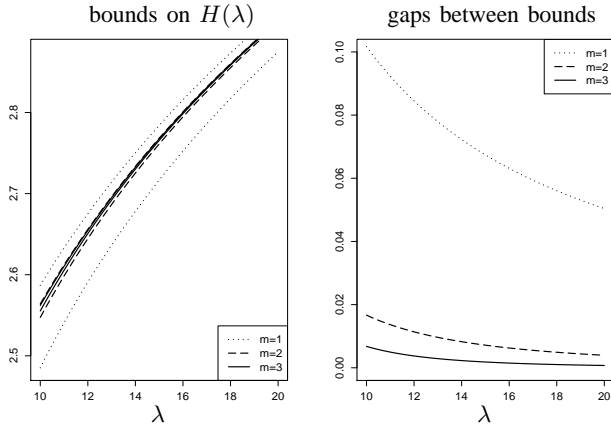| k	$b(m,k)$			
1	1/6	−1/12	−1/12	−1/12
2		5/24	−1/24	−1/24
3		1/60	103/180	−19/360
4			13/40	201/80
5			1/210	12367/2520
6				571/1008
7				1/504

Fig. 1. Bounds on $H(\lambda)$ (left) and differences between upper and lower bounds (right) given by Theorem 2 with $m = 1,2,3$.

• from ≈ 0.0068 to ≈ 0.00074 with $m = 3$.

In short, for moderate λ, the bounds are already quite accurate with m as small as 3, and, as expected, the accuracy improves as λ increases.

B. Exact Formulae and Sharp Bounds for $D(n,p)$

While bounds on Poisson convergence are often stated in terms of the total variation distance ([4], [2]), those based on the relative entropy can also be quite effective ([21], [15]). As a discrepancy measure, relative entropy occurs naturally in contexts such as hypothesis testing. In this section we consider $D(n,p) \equiv D(B_{n,p} \| N_{np})$, and study its higher-order asymptotic behavior. Our main result, Theorem 4, may be regarded as a refinement of that of Harremoës and Ruzankin [15]. As a by-product, we also obtain bounds on the entropy of the binomial that parallel Theorem 2.

Analogous to Theorem 1, we have the following exact expansion for $D(n,p)$ as a function of p.

Theorem 3: Fix $n \in \mathbf{N}$. For $p \in [0,1]$ we have

$$D(n,p) = n(p + q\log q) + \sum_{k=2}^{n} \binom{n}{k} \tilde{c}(k)p^k,$$

where

$$\tilde{c}(k) = \sum_{j=0}^{k-1} (-1)^{k-1-j} \binom{k-1}{j} \log(n-j), \ k = 2, \dots, n.$$

In the following result, the kth central moment of $B_{n,p}$,

$$\mu_k(n,p) \equiv E[(B_{n,p} - np)^k],$$

plays an important role. The first few values of $\mu_k(n,p)$ are

$$\mu_0(n,p) = 1, \ \mu_1(n,p) = 0, \ \mu_2(n,p) = pqn,$$

$$\mu_3(n,p) = (q-p)pqn, \ \mu_4(n,p) = 3(pqn)^2 + (1 - 6pq)pqn.$$

We write $q \equiv 1 - p$ throughout.

Analogous to Theorem 2, Theorem 4 gives bounds on $D(n,p)$ that are effective for large n.

Theorem 4: For $n, m \in \mathbf{N}$ and $p \in (0,1)$, we have

$$0 \le D(n,p) + \frac{p + \log q}{2} - \tilde{\beta}_m(n,p) \le \tilde{r}_m(n,p),$$

where

$$\tilde{\beta}_m(n,p) = n \int_q^1 \sum_{j=3}^{2m+1} \frac{(-1)^j \mu_j(n,s)}{j(j-1)(ns)^j} \, ds = \sum_{k=1}^{2m-1} \frac{\tilde{b}(m,k;p)}{n^k}$$

and

$$\tilde{r}_m(n,p) = n \int_q^1 \frac{\mu_{2m+2}(n,s)}{(2m+1)(ns)^{2m+2}} \, ds = \sum_{k=m}^{2m} \frac{\tilde{a}(m,k;p)}{n^k}.$$

The integrals that define $\tilde{\beta}_m(n,p)$ and $\tilde{r}_m(n,p)$ are easy to calculate, because $\mu_j(n,s)$ is a polynomial in s. We obtain the coefficients $\tilde{b}(m,k;p)$ and $\tilde{a}(m,k;p)$ after integrating and assembling the results in powers of n^{-1}. For example, we have ($m = 1$)

$$\tilde{b}(1,1;p) = -\frac{\log q}{2} + \frac{q}{3} - \frac{1}{6q} - \frac{1}{6},$$

$$\tilde{a}(1,1;p) = 2\log q - q + \frac{1}{q},$$

$$\tilde{a}(1,2;p) = -4\log q + 2q - \frac{7}{3q} + \frac{1}{6q^2} + \frac{1}{6},$$

and ($m = 2$)

$$\tilde{b}(2,1;p) = \frac{p^2}{12q}, \qquad (6)$$

$$\tilde{b}(2,2;p) = \frac{3}{2}\log q - \frac{1}{2}q + \frac{17}{12q} - \frac{5}{24q^2} - \frac{17}{24},$$

$$\tilde{b}(2,3;p) = -3\log q + \frac{6}{5}q - \frac{5}{2q} + \frac{3}{8q^2} - \frac{1}{60q^3} + \frac{113}{120},$$

$$\tilde{a}(2,2;p) = -9\log q + 3q - \frac{9}{q} + \frac{3}{2q^2} + \frac{9}{2},$$

$$\tilde{a}(2,3;p) = 78\log q - 26q + \frac{83}{q} - \frac{18}{q^2} + \frac{5}{3q^3} - \frac{122}{3},$$

$$\tilde{a}(2,4;p) = -72\log q + 24q - \frac{78}{q} + \frac{18}{q^2} - \frac{31}{15q^3}$$
$$+ \frac{1}{20q^4} + \frac{2281}{60}.$$

As in Theorem 2, we emphasize that, for each $m \ge 2$, the constants $\tilde{b}(m,k;p)$, $1 \le k \le m-1$, are exact in the full

asymptotic expansion of $D(n, p)$ (for fixed p as $n \to \infty$). In particular, from (6) we get

$$D(n, p) = -\frac{p + \log q}{2} + \frac{p^2}{12qn} + O(n^{-2}),$$

for fixed $p \in (0, 1)$. We can also estimate the rate at which $D(n, \lambda/n)$ decreases to zero, for fixed $\lambda > 0$ as $n \to \infty$, which corresponds to the usual binomial-to-Poisson convergence. Indeed, setting $m = 1$ and $p = \lambda/n$, Theorem 4 yields, after routine calculations,

$$D(n, \lambda/n) = \frac{\lambda^2}{4n^2} + O(n^{-3}),$$

which can be further refined by using larger m.

Such results are related to those of Harremoës and Ruzankin [15], who give several bounds on $D(n, p)$ after detailed analyses on inequalities involving Stirling numbers. Theorem 4 may be regarded as a refinement in that, by letting m be arbitrary, it can give a full asymptotic expansion of $D(n, p)$, with computable bounds. Our derivation is also simpler (see Section IV).

Let us denote the entropy of the binomial by $H(n, p) = H(B_{n,p})$. In the special case $p = 1/2$, $H(n, p)$ appears as the sum capacity of a noiseless n-user binary adder channel as analyzed by Chang and Weldon [7], who also provide simple bounds on $H(n, p)$. For general p, asymptotic expansions for $H(n, p)$ have been obtained by Jacquet and Szpankowski [17] and Knessl [20] (see also Flajolet [11]). We present sharp bounds on $H(n, p)$ that complement such expansions. As it turns out, due to an elementary identity (see (29) below), bounds on $D(n, p)$ obtained in Theorem 4 translate directly to those on $H(n, p)$, thus simplifying our analysis.

Corollary 1: Let $n, m \in \mathbf{N}$ and $p \in (0, 1)$. Define

$$\tilde{H}(n, p) = H(n, p) - \log n! + n \log n - n - \frac{1 + \log(pq)}{2}.$$

Then

$$0 \le -\tilde{H}(n, p) - \sum_{k=1}^{2m-1} \frac{\tilde{b}(m, k; p) + \tilde{b}(m, k; q)}{n^k}$$

$$\le \sum_{k=m}^{2m} \frac{\tilde{a}(m, k; p) + \tilde{a}(m, k; q)}{n^k},$$

where $\tilde{a}(m, k; p)$ and $\tilde{b}(m, k; p)$ are defined as in Theorem 4.

Bounds on $H(n, p)$ can also be expressed in terms of $\log n$ and n^{-k}, $k = 1, 2, \ldots$, via familiar bounds on $\log n!$. For example, taking $m = 1$ in Corollary 1, and using (see [1], 6.1.42)

$$\frac{1}{12n} - \frac{1}{360n^3} < \log n! - n \log n + n - \frac{1}{2} \log(2\pi n) < \frac{1}{12n},$$

we get

$$\frac{C_1}{n} + \frac{C_2}{n^2} + \frac{C_3}{n^3} < H(n, p) - \frac{1}{2} \log(2\pi npq) - \frac{1}{2} < \frac{C_4}{n}, \quad (7)$$

where

$$C_1 = \frac{13}{12} - \frac{3 \log(pq)}{2} - \frac{5}{6pq},$$

$$C_2 = -\frac{7}{3} + 4 \log(pq) + \frac{8}{3pq} - \frac{1}{6(pq)^2},$$

$$C_3 = -\frac{1}{360},$$

$$C_4 = \frac{1}{12} + \frac{\log(pq)}{2} + \frac{1}{6pq}.$$

To relate (7) to the Poisson case, i.e., Theorem 2, let us fix $\lambda > 0$ and set $p = \lambda/n$. Then, as $n \to \infty$, we have $H(n, p) \to H(\lambda)$, and in the limit (7) becomes

$$-\frac{5}{6\lambda} - \frac{1}{6\lambda^2} \le H(\lambda) - \frac{1}{2} \log(2\pi\lambda) - \frac{1}{2} \le \frac{1}{6\lambda},$$

which is precisely Theorem 2 with $m = 1$. In general ($m \ge 1$), Theorem 2 may be seen as a limiting case of Corollary 1.

As before, for each $m \ge 2$, the coefficients $\tilde{b}(m, k; p) + \tilde{b}(m, k; q)$, $k = 1, \ldots, m - 1$, are exact in the asymptotic expansion of $\tilde{H}(n, p)$ (for fixed p as $n \to \infty$). For large n, we may choose larger m to improve the accuracy.

III. Proofs of Theorems 1 and 2

A. Preliminaries

Given a function $\phi : \mathbf{R} \to \mathbf{R}$, the mth forward difference of ϕ is defined recursively by

$$\Delta^0 \phi(x) = \phi(x), \quad \Delta^{m+1} \phi(x) = \Delta^m \phi(x + 1) - \Delta^m \phi(x),$$

for any $x \in \mathbf{R}$ and $m \in \mathbf{Z}_+$. Equivalently, we have

$$\Delta^m \phi(x) = \sum_{j=0}^{m} (-1)^{m-j} \binom{m}{j} \phi(x + j), \; x \in \mathbf{R}, \; m \in \mathbf{Z}_+. \quad (8)$$

(This definition extends to functions defined only on $\mathbf{Z}_+$.) As usual, $\phi^{(m)}$ stands for the mth derivative of ϕ. If ϕ is infinitely differentiable, then

$$\Delta^m \phi(x) = E[\phi^{(m)}(x + S_m)], \quad (9)$$

where $S_m = U_1 + \cdots + U_m$ and $(U_k)_{k \ge 1}$ is a sequence of independent and identically distributed (i.i.d.) random variables having the uniform distribution on $[0, 1]$ (cf. [3], Eqn. (2.7)).

On the other hand, two special properties of the Poisson distribution are

$$\lambda E[\phi(N_\lambda + 1)] = E[N_\lambda \phi(N_\lambda)], \quad \lambda \ge 0, \quad (10)$$

and (see [3], Theorem 2.1, for example)

$$\frac{\mathrm{d}^m E[\phi(N_\lambda)]}{\mathrm{d}\lambda^m} = E[\Delta^m \phi(N_\lambda)], \quad \lambda \ge 0, \; m \in \mathbf{Z}_+. \quad (11)$$

In both (10) and (11), $\phi : \mathbf{Z}_+ \to \mathbf{R}$ is an arbitrary function for which the relevant expectations exist. From (11) we have the Taylor formula

$$E[\phi(N_\lambda)] = \sum_{k=0}^{m} \frac{\Delta^k \phi(0)}{k!} \lambda^k$$

$$+ \frac{1}{m!} \int_0^\lambda (\lambda - u)^m E[\Delta^{m+1} \phi(N_u)] \, \mathrm{d}u, \quad (12)$$

for any $\lambda \ge 0$ and $m \in \mathbf{Z}_+$.

B. Proof of Theorem 1

Let $\lambda \geq 0$. We have

$$H(\lambda) = \lambda - \lambda \log \lambda + E[\log N_\lambda!] \qquad (13)$$

by definition. Applying (12) to the function $\phi(k) = \log k!$, $k \in \mathbf{Z}_+$, we obtain

$$H(\lambda) + \lambda \log \lambda - \lambda = \sum_{k=2}^{m} \frac{\Delta^k \phi(0)}{k!} \lambda^k$$
$$+ \frac{1}{m!} \int_0^\lambda (\lambda - u)^m E[\Delta^{m+1}\phi(N_u)]\, du, \qquad (14)$$

for any $m = 2, 3, \ldots$, since $\Delta^0 \phi(0) = \Delta^1 \phi(0) = 0$.

On the other hand, letting $g(x) = \log(x+1)$, $x \geq 0$, and using (8) and (9), we can write

$$\Delta^{m+1}\phi(i) = \Delta^m g(i)$$
$$= (-1)^{m-1} E\left[\frac{(m-1)!}{(i+1+S_m)^m}\right] \qquad (15)$$
$$= \sum_{j=0}^{m} (-1)^{m-j} \binom{m}{j} \log(i+1+j),$$

where $i \in \mathbf{Z}_+$ and $m = 1, 2, \ldots$. Therefore,

$$\Delta^k \phi(0) = \sum_{j=0}^{k-1} (-1)^{k-1-j} \binom{k-1}{j} \log(j+1) = c(k), \quad (16)$$

for $k = 2, 3, \ldots$. The conclusion follows from (14) and (16) by noting that, based on (15), the integral in (14) alternates in sign for $m = 1, 2, \ldots$.

C. Proof of Theorem 2

Some auxiliary results are needed in the proof of Theorem 2. In Lemmas 1 and 2 we denote

$$\tilde{H}(\lambda) = H(\lambda) - \frac{1}{2}\log(2\pi\lambda) - \frac{1}{2}, \quad \lambda > 0, \qquad (17)$$

for convenience.

Lemma 1: We have $\tilde{H}(\lambda) = O(\lambda^{-1})$, as $\lambda \to \infty$.

Proof: See [10] or [20]. ∎

Lemma 2 expresses the quantity of interest in terms of an easier quantity, $E[\log(N_s + 1)]$.

Lemma 2: For $\lambda > 0$ we have

$$\tilde{H}(\lambda) = \int_\lambda^\infty \left(\frac{1}{2s} - E\left[\log \frac{N_s+1}{s}\right]\right)ds.$$

Proof: Recalling (13) and using (11) with $\phi(k) = \log k!$, $k \in \mathbf{Z}_+$ and $m = 1$, we obtain from (17)

$$\tilde{H}'(\lambda) = E[\log(N_\lambda + 1)] - \log \lambda - \frac{1}{2\lambda}.$$

By Lemma 1, $\tilde{H}(\infty) = 0$, and the claim follows. ∎

The next result presents sharp bounds on $E[\log(N_s + 1)]$.

Proposition 1: For $s > 0$ and $m \in \mathbf{N}$, we have

$$0 \leq E\left[\log \frac{N_s+1}{s}\right] - \sum_{k=2}^{2m+1} \frac{(-1)^k \mu_k(s)}{k(k-1)s^k} \leq \frac{\mu_{2m+2}(s)}{(2m+1)s^{2m+2}}.$$

Proof: By (10) we have

$$E\left[\log \frac{N_s+1}{s}\right] = \frac{1}{s}\left(E[N_s \log N_s] - s \log s\right). \qquad (18)$$

Taking into account that $E[N_s] = s$, the lower bound follows from (18) and the inequality (upon letting $x = N_s$)

$$x \log x - s \log s \geq (\log s + 1)(x - s) + \sum_{k=2}^{2m+1} \frac{(-1)^k (x-s)^k}{k(k-1)s^{k-1}}, \qquad (19)$$

which holds for $x \geq 0$, $s > 0$.

On the other hand, using the inequality

$$\log(1+x) \leq \sum_{k=1}^{2m+1} \frac{(-1)^{k-1}}{k} x^k, \quad x > -1, \qquad (20)$$

with $x = (N_s - s + 1)/s$, we obtain

$$E\left[\log \frac{N_s+1}{s}\right] \leq \sum_{k=1}^{2m+1} \frac{(-1)^{k-1} E[(N_s - s + 1)^k]}{ks^k}. \qquad (21)$$

Again by (10), we have

$$sE[(N_s - s + 1)^k] = E[N_s(N_s - s)^k] = \mu_{k+1}(s) + s\mu_k(s).$$

Hence the right-hand side of (21) is equal to

$$\frac{\mu_{2m+2}(s)}{(2m+1)s^{2m+2}} + \sum_{k=2}^{2m+1} \frac{(-1)^k \mu_k(s)}{k(k-1)s^k},$$

which proves the upper bound. ∎

Remark. The quantity $E[\log(N_s + 1)]$ already appears as Example 2 of Entry 10 in Chapter 3 of Ramanujan's second notebook ([5]). Proposition 1 gives a finite expansion of $E[\log(N_s + 1)]$ with an explicit upper bound of the order of $s^{-(m+1)}$. See [31] for related work on this particular entry of Ramanujan.

Proof of Theorem 2: The claim follows directly from Lemma 2 and Proposition 1, upon noting $\mu_2(s) = s$. ∎

IV. PROOFS OF THEOREMS 3 AND 4 AND COROLLARY 1

A. Preliminaries

Let us recall two special properties of the binomial variable $B_{n,p}$. For $n \in \mathbf{N}$, $p \in [0,1]$, and any function $\phi : \{0, 1, \ldots, n\} \to \mathbf{R}$, we have

$$npE[\phi(B_{n-1,p} + 1)] = E[B_{n,p}\phi(B_{n,p})], \qquad (22)$$

as well as (see [2], Theorem 1, for example)

$$\frac{d^k E[\phi(B_{n,p})]}{dp^k} = k!\binom{n}{k} E[\Delta^k \phi(B_{n-k,p})], \quad k = 0, 1, \ldots, n; \qquad (23)$$

$$\frac{d^k E[\phi(B_{n,p})]}{dp^k} = 0, \quad k = n+1, n+2, \ldots.$$

From (23) we obtain the Taylor formula

$$E[\phi(B_{n,p})] = \sum_{k=0}^{n} \binom{n}{k} E[\Delta^k \phi(B_{n-k,t})](p-t)^k, \qquad (24)$$

where $p, t \in [0,1]$.

B. Proof of Theorem 3

By direct calculation

$$D(n,p) = n(p + q\log q) - np\log n + E\left[\log\frac{n!}{(n-B_{n,p})!}\right]. \quad (25)$$

Using (24) with $\phi(l) = \log(n!/(n-l)!)$, $l = 0,\ldots,n$ and $t = 0$, we see that the expectation in (25) is equal to

$$E[\phi(B_{n,p})] = \sum_{k=0}^{n}\binom{n}{k}\Delta^k\phi(0)p^k, \quad p \in [0,1]. \quad (26)$$

Denote by $g(l) = \log(n-l)$, $l = 0,1,\ldots,n-1$. Observe that

$$\Delta^k\phi = \Delta^{k-1}g, \ k \in \mathbf{N}, \quad \Delta^0\phi(0) = 0, \quad \Delta^1\phi(0) = \log n.$$

Therefore, we have from (8)

$$\Delta^k\phi(0) = \Delta^{k-1}g(0)$$
$$= \sum_{j=0}^{k-1}(-1)^{k-1-j}\binom{k-1}{j}\log(n-j) = \tilde{c}(k), \quad (27)$$

for $k = 2,\ldots,n$. The claim follows from (25), (26), and (27). ∎

C. Proof of Theorem 4

The following integral representation is crucial in the proof of Theorem 4.

Lemma 3: For any $p \in [0,1]$ we have

$$D(n,p) = n\int_q^1 E\left[\log\frac{B_{n-1,s}+1}{ns}\right]\mathrm{d}s.$$

Proof: Differentiating (25) once and applying (23) with $k = 1$ and $\phi(l) = \log(n!/(n-l)!)$, we get

$$\frac{\mathrm{d}D(n,p)}{\mathrm{d}p} = nE[\log(n - B_{n-1,p})] - n\log(nq).$$

Since $D(n,0) = 0$, we have

$$D(n,p) = n\int_0^p (E[\log(n - B_{n-1,t})] - \log(n(1-t)))\,\mathrm{d}t.$$

Note that $n - B_{n-1,t}$ and $B_{n-1,1-t} + 1$ have the same distribution. The claim follows by a change of variables $s = 1 - t$. ∎

Proposition 2 gives upper and lower bounds on the key quantity $E[\log((B_{n-1,s} + 1)/(ns))]$. This parallels Proposition 1.

Proposition 2: For any $m \in \mathbf{N}$ and $s \in (0,1)$ we have

$$0 \le E\left[\log\frac{B_{n-1,s}+1}{ns}\right] - \sum_{k=2}^{2m+1}\frac{(-1)^k\mu_k(n,s)}{k(k-1)(ns)^k}$$
$$\le \frac{\mu_{2m+2}(n,s)}{(2m+1)(ns)^{2m+2}}.$$

Proof: By (22) we have

$$nsE[\log(B_{n-1,s}+1)] = E[B_{n,s}\log(B_{n,s})].$$

The lower bound follows from this and (19).

As in the proof of Lemma 2, using (20) with

$$x = \frac{B_{n-1,s}+1-ns}{ns},$$

we obtain

$$E\left[\log\frac{B_{n-1,s}+1}{ns}\right] \le \sum_{k=1}^{2m+1}\frac{(-1)^{k-1}E[(B_{n-1,s}-ns+1)^k]}{k(ns)^k}. \quad (28)$$

Again by (22), we have

$$nsE[(B_{n-1,s}-ns+1)^k] = E[B_{n,s}(B_{n,s}-ns)^k]$$
$$= \mu_{k+1}(n,s) + ns\mu_k(n,s).$$

This, in conjunction with (28), proves the upper bound. ∎

Proof of Theorem 4: Note that

$$n\int_q^1\frac{\mu_2(n,s)}{2(ns)^2}\,\mathrm{d}s = -\frac{p + \log q}{2}.$$

The claim then follows from Lemma 3 and Proposition 2. ∎

D. Proof of Corollary 1

From the definitions we have

$$H(n,p) = \log n! - n\log n + n - D(n,p) - D(n,q), \quad (29)$$

where $q \equiv 1 - p$ as before. Hence, Corollary 1 is an immediate consequence of (29) and Theorem 4. ∎

V. Discussion

We have obtained asymptotically sharp and readily computable bounds on the Poisson entropy function $H(\lambda)$. The method also handles the entropy of the binomial, and the relative entropy $D(n,p)$ between the binomial(n,p) and Poisson(np) distributions, yielding full asymptotic expansions with explicit constants. While some results are of theoretical interest, bounds on the entropy are intended to aid channel capacity calculations for discrete-time Poisson channels, for example.

Besides entropy calculations, the method also extends to quantities such as the fractional moments of these familiar distributions. An example is $E[\sqrt{N_\lambda}]$, which appears as Example 1, Entry 10, in Chapter 3 of Ramanujan's notebook ([5]). Analogous to Lemma 2, a double inequality can be obtained for $E[\sqrt{N_\lambda}]$ (details omitted). This is of some statistical interest because it gives accurate bounds on the bias associated with the square root transformation, which is variance-stabilizing for the Poisson.

For theoretical considerations, the Taylor formulae, integral representations, and related results can also help to establish interesting monotonicity properties of quantities such as $H(\lambda)$. For example, it can be shown that $H'(\lambda)$ is a completely monotonic function of λ, i.e., $(-1)^{k-1}H^{(k)}(\lambda) \ge 0$ for all $k \ge 1$. It appears that many entropy-like functions associated with classical distributions are completely monotonic ([32]). One conjecture ([33], Conjecture 1) states that $D(n,\lambda/n)$, $n > \lambda$, is completely monotonic in n for fixed $\lambda > 0$. The method of this work may prove useful toward resolving such conjectures.

REFERENCES

[1] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions*, Dover, New York, 1964.
[2] J. A. Adell and J. M. Anoz, "Signed binomial approximation of binomial mixtures via differential calculus for linear operators," *J. Statist. Plann. Inference,* vol. 138, no. 12, pp. 3687–3695, 2008.
[3] J. A. Adell and A. Lekuona, "Sharp estimates in signed Poisson approximation of Poisson mixtures," *Bernoulli,* vol. 11, no. 1, 47–65, 2005.
[4] A. D. Barbour, L. Holst, and S. Janson, *Poisson Approximation, Oxford Studies in Probability*, vol. 2, Clarendon Press, Oxford, 1992.
[5] B. C. Berndt, *Ramanujan's Notebooks, Part I*, Springer-Verlag, New York, 1985.
[6] D. Brady and S.Verdú, "The asymptotic capacity of the direct detection photon channel with a bandwidth constraint," in *Proc. 28th Allerton Conf. Communication, Control and Computing*, Allerton House, Monticello, IL, Oct. 1990, pp. 691-700.
[7] S. Chang and E. Weldon, "Coding for T-user multiple-access channels," *IEEE Trans. Inf. Theory*, vol. IT-25, pp. 684-691, 1979.
[8] T. Cover and J. Thomas, *Elements of Information Theory*, 2nd ed., New York: Wiley, 2006.
[9] A. D'yachkov and P. Vilenkin, "Asymptotics of the Shannon and Rényi entropies for sums of independent random variables," in *Proc. ISIT98* (MIT, Cambridge, MA, 1998), p. 376.
[10] R. J. Evans, J. Boersma, N. M. Blachman and A. A. Jagers, "The entropy of a Poisson distribution: problem 87-6," *SIAM Review*, vol. 30, pp. 314–317, 1988.
[11] P. Flajolet, "Singularity analysis and asymptotics of Bernoulli sums," *Theor. Comput. Sci.*, vol. 275, pp. 371-387, 1999.
[12] M. R. Frey, "Information capacity of the Poisson channel," *IEEE Trans. Inf. Theory*, vol. 37, no. 2, pp. 244-256, Mar. 1991.
[13] D. Guo, S. Shamai and S. Verdú, "Mutual information and conditional mean estimation in Poisson channels," *IEEE Trans. Inf. Theory*, vol. 54, no. 5, pp. 1837–1849, May 2008.
[14] P. Harremoës, "Binomial and Poisson distributions as maximum entropy distributions," *IEEE Trans. Inf. Theory*, vol. 47, no. 5, pp. 2039–2041, July 2001.
[15] P. Harremoës and P. S. Ruzankin, "Rate of convergence to Poisson law in terms of information divergence," *IEEE Trans. Inf. Theory*, vol. 50, no. 9, pp. 2145–2149, 2004.
[16] P. Jacquet and W. Szpankowski, "Analytical depoissonization and its applications," *Theoret. Comput. Sci.*, vol. 201, pp. 1–62, 1998.
[17] P. Jacquet and W. Szpankowski, "Entropy computations via analytic depoissonization," *IEEE Trans. Inf. Theory*, vol. 45, pp. 1072–1081, 1999.
[18] O. Johnson, "Log-concavity and the maximum entropy property of the Poisson distribution," *Stochastic Processes and their Applications*, vol. 117, no. 6, pp. 791–802, Jun. 2007.
[19] N. L. Johnson, A. W. Kemp and S. Kotz, *Univariate Discrete Distributions*, 3rd ed., Wiley & Sons, Hoboken, NJ, 2005.
[20] C. Knessl, "Integral representations and asymptotic expansions for Shannon and Renyi entropies," *Appl. Math. Lett.*, vol. 11, pp. 69-74, 1998.
[21] I. Kontoyiannis, P. Harremoës, and O. T. Johnson, "Entropy and the law of small numbers," *IEEE Trans. Inf. Theory*, vol. 51, no. 2, pp. 466–472, Feb. 2005.
[22] A. Lapidoth and S. M. Moser, "On the capacity of the discrete-time Poisson channel," *IEEE Trans. Inf. Theory*, vol. 55, no. 1, pp. 303–322, 2009.
[23] M. Madiman, O. Johnson and I. Kontoyiannis, "Fisher information, compound Poisson approximation and the Poisson channel," in *Proc. IEEE International Symposium on Information Theory*, Nice, France, Jun. 2007.
[24] A. Martinez, "Spectral efficiency of optical direct detection," *J. Opt. Soc. America B*, vol. 24, no. 4, pp. 739-749, Apr. 2007.
[25] P. Mateev, "The entropy of the multinomial distribution," *Teor. Verojatnost. i Primenen.*, vol. 23, no. 1, pp. 196–198, 1978.
[26] S. Shamai (Shitz), "Capacity of a pulse amplitude modulated direct detection photon channel," in *Proc. Inst. Elec. Eng.*, vol. 137, no. 6, pp. 424-430, Dec. 1990, part I (Communications, Speech and Vision).
[27] S. Shamai (Shitz) and A. Lapidoth, "Bounds on the capacity of a spectrally constrained Poisson channel," *IEEE Trans. Inf. Theory*, vol. 39, no. 1, pp. 19-29, Jan. 1993.
[28] L. A. Shepp and I. Olkin, "Entropy of the sum of independent Bernoulli random variables and of the multinomial distribution," in *Contributions to Probability*, pp. 201–206, Academic Press, New York, 1981.
[29] F. Topsøe, "Maximum entropy versus minimum risk and applications to some classical discrete distributions," *IEEE Trans. Inf. Theory*, vol. 48, no. 8, pp. 2368–2376, Aug. 2002.
[30] Y. Yu, "On the maximum entropy properties of the binomial distribution," *IEEE Trans. Inf. Theory*, vol. 54, pp. 3351–3353, 2008.
[31] Y. Yu, "On an asymptotic series of Ramanujan," *Ramanujan J.*, vol. 20, pp. 179–188, 2009.
[32] Y. Yu, "Complete monotonicity of the entropy in the central limit theorem for gamma and inverse Gaussian distributions," *Stat. Prob. Lett.*, vol. 79, pp. 270–274, 2009.
[33] Y. Yu, "Monotonic convergence in an information-theoretic law of small numbers," *IEEE Trans. Inf. Theory*, vol. 55, pp. 5412–5422, 2009.

José A. Adell received the B.S. degree in mathematics from Zaragoza University, Spain, in 1980, and the Ph.D. degree in mathematics from the Basque Country University, Spain, in 1983. During 1981–1990, he was Assistant Professor in the Department of Mathematics at the Basque Country University. During 1991–2009, he has been Associate Professor in the Department of Statistics at Zaragoza University.

Alberto Lekuona received the B.S. degree in mathematics from Zaragoza University, Spain, in 1985, and the Ph.D. degree in mathematics from Zaragoza University, in 1996. During 1987–2000, he was Assistant Professor in the Department of Statistics at Zaragoza University, and Associate Professor at the same place in the period 2001–2009.

Yaming Yu (M'08) received the B.S. degree in mathematics from Beijing University, P.R. China, in 1999, and the Ph.D. degree in statistics from Harvard University, in 2005. Since 2005 he has been an Assistant Professor in the Department of Statistics at the University of California, Irvine.